November 21 , 2000

CHEMICAL IDENTIFICATION DATA STANDARD BUSINESS RULES

United States Environmental Protection Agency
Office of Environmental Information
1200 Pennsylvania Avenue, NW.
Washington, DC 20460

CONTENTS

November 21 , 2000

BUSINESS RULES FOR THE CHEMICAL IDENTIFICATION DATA STANDARD

1. THE STANDARD

a. This standard specifies the key data elements necessary to constitute unambiguous identification of a chemical substance or a chemical grouping of interest to the Environmental Protection Agency (EPA).

b. The *Interim Data Standard for Chemical Identification* specifies the key data elements necessary to constitute a viable chemical substance record. It is based on the Chemical Abstracts Service (CAS) registry numbering standard but also includes additional identification elements. This standard requires that a complete record be centrally created and maintained. The key elements are:

- Chemical Abstracts Service number.
- Chemical Abstracts Service name (9th Collective Index Name).
- EPA [Registry] name.
- Unique record number.

The central record of all chemical substances of interest to the Agency will not display the unique record number for chemicals with CAS numbers and will also maintain optional identification data to better meet the Agency's need for unambiguous identification, to the extent possible.

c. The purpose of this standard is to:

1) Provide a common and consistent way to identify and represent chemical substances across the Agency.

2) Provide EPA with a unique, unambiguous, chemically correct common name for each chemical substance and chemical grouping in which the Agency has an interest.

3) Provide a way to reference data about chemical substances across EPA systems and provide a basis for searching for chemicals in these systems in an automated way.

2. DEFINITIONS

a. A *chemical substance* is any material that can be categorized and defined for the purpose of this standard as follows:

1) A *single fully-defined chemical substance* consists of a single element or compound uniquely identified by the number and nature of the atoms present, the bonds which connect each of the atoms, the nature of the bonds, and the spatial orientation, including the following:

a) Specification of atomic number, atomic weight for isotopes or tagged atoms (e.g., urea-14C, acetone-d6), and oxidation state (e.g., iron(II) chloride or ferrous chloride).

b) Bonding and type of bond that describes and differentiates straight and branched isomers (e.g., tert-butylacetic acid, iso-propylalcohol).

c) Bridgehead positions (e.g., bicyclo[2.2.1]hept-2-ene norbornene).

d) Substituent locations (e.g., o-xylene or 1,2-xylene; 1-octene).

e) Saturation/unsaturation and the associated location(s) (e.g., 1(3H)-isobenzofuranone).

f) Spatial orientation that unambiguously identifies structural or stereoisomers (e.g., cis-1,2-dichloroethene or (Z)-1,2-dichloroethene; (R)-2-chlorobutane or d-2-chlorobutane).

2) A *chemical species* is a component of a single, fully-defined chemical substance such as an ion, functional group, or biological available form (e.g., nitrate nitrogen, phosphate).

3) A chemical substance of *known composition* comprises two or more single fully-defined chemical substances that are always present in the same precisely-defined ratio (e.g., formulated mixtures, fixed-ratio salts, fully-defined isomer mixtures).

4) A chemical substance of *variable composition* is one for which all possible constituents are known single fully-defined chemical substances, but it is not known which constituents are present or, if present, in what ratio (e.g., variable isomer mixtures, undefined-ratio salts, PCB mixtures [i.e., Aroclors]).

5) A chemical substance of *unknown composition* is one for which one or more constituents is unknown (e.g., complex reaction products, naturally occurring substances, biologicals). Some chemical substances of this type may be described in terms of the original constituents of a reaction; others in terms of the product composition.

6) A *generic-CBI* chemical substance is one for which the Agency is not permitted to release an exact identity due to a claim that the exact identity is Confidential Business Information (CBI). Initially, these will not be included in the Chemical Registry.

b. A *chemical grouping* is defined for the purpose of this standard, as a number of individual chemical substances considered together as being related in some way and for which the Agency has an interest, including the following types of groupings:

1) A *chemical class* is a set of individual substances grouped together due to chemical similarity (e.g., lead and lead compounds).

2) A *chemical category* is a set of individual substances grouped together due to a similarity other than chemical (e.g., surfactants, carpet adhesives).

c. The EPA *Environmental Data Registry* (EDR) is the central repository and reference tool for Agency data standards, data elements, and other objects, such as business rules.

d. The *Chemical Registry* (CR) is the central collection of identification data for chemical substances and chemical groupings of interest to EPA.

e. The *Chemical Registry Database* is the collection of tables where the chemical identification data is stored in a database management system.

f. The *Chemical Registry System* (CRS) is the central Agency tool for search and retrieval of chemical identification data, for system integration, and for sharing identification data about chemical substances and chemical groupings of interest to EPA and external stakeholders and partners.

g. The *Chemical Identification Standing Committee* is a management committee composed of representatives from each EPA office that manages chemical information.

h. The *EPA Chemical Nomenclature Committee* is a small group of EPA chemists/ technical staff with expertise in chemical nomenclature as it applies to chemical substances and groupings of concern to EPA.

i. The *Data Steward for the CR* is the person or organization to whom is delegated the responsibility for managing the data content of the CR.

j. The *Data Standard Steward for Chemical Substance Identification* is the person or organization to whom is delegated the responsibility for managing the data resources for the Chemical Identification Data Standard.

3. APPLICABILITY

a. This standard is applicable to the 13 REI-designated EPA national program systems.

b. This standard is applicable to EPA documents and publications, databases, regulations, rules, and Federal Register Notices created by EPA staff and contractors.

c. This standard is applicable to State and Tribal partners and other stakeholders in the following contexts:

1) The standard will support the coordination and exchange of chemistry identification data where EPA and its partners have agreed to do so.

2) This standard does not change EPA-partner obligations that are embodied in statutes, regulations, delegation agreements, and other individual EPA-partner agreements.

3) This standard allows trading partners and data managers to establish rules about the specific, standard chemical identification data elements they need to manage, as their business needs dictate.

d. This standard is applicable to new data collections and information systems and to reengineered information systems.

e. This standard does not apply to retired systems not presently in use or to documents and publications that were created prior to the adoption of this standard.

4. DATA REQUIREMENTS

a. The minimum data storage requirement for chemical substances and chemical groupings in the 13 REI-designated national program systems and in all newly created or reengineered EPA data systems is the following:

1) EPA Chemical Internal Tracking Number (DE 20040:1) – the unique record number assigned to all chemical substances and chemical groupings for internal tracking within EPA systems. This data element is mandatory. It is an electronic key that facilitates data exchange with the CRS and other EPA databases. This identification number must be obtained from the Chemical Registry.

2) Other identifiers and names may also be stored in the 13 REI-designated EPA national program systems and in newly created or reengineered EPA data systems at the discretion of the responsible EPA program. These other identifiers and names shall be obtained from and validated by the Chemical Registry.

b. The central EPA Chemical Registry System (CRS) must store the following mandatory data element for all chemical substances and chemical groupings of interest to the Agency:

1) EPA Chemical Internal Tracking Number (DE 20040:1) – the unique record number assigned to all chemical substances and chemical groupings for internal tracking within EPA systems. This mandatory data element is an electronic key used to facilitate data exchange with the CRS and other EPA databases. The EPA Chemical Internal Tracking Number is not intended for public use.

c. The CRS must store the following standard chemical identification data elements where the data exists or can be obtained:

1) Chemical Abstracts Service Registry Number (DE 6165:1) – the unique number assigned by Chemical Abstracts Service (CAS) to a chemical substance is mandatory where that number exists or can be assigned.

2) The Chemical Substance Systematic Name (DE 5799:1) – the name assigned to a chemical substance that describes it in terms of its molecular composition is mandatory where that name exists or can be assigned. Note: The systematic name preferred by EPA is the index name formulated according to the nomenclature rules set forth for the CAS 9th Collective Indexing Period.

3) The EPA Chemical Registry Name (DE 5801:1) – the name EPA has selected as the name to be commonly used by EPA in referring to a chemical substance. The chemical registry name cannot be assigned to all chemical substances of interest to the Agency. The name, however, is mandatory for all chemical groupings and for chemical substances where CAS systematic names do not exist and cannot be assigned.

4) The EPA Chemical Identifier (DE 19994:1) – the identifier to be created and placed in the Chemical Registry for each chemical substance or chemical grouping in the Chemical Registry for which a CAS Registry Number does not exist and cannot be assigned. This identifier is appropriate to be used for data transfer.

d. Other data elements that provide additional identification information about a chemical substance or chemical grouping will also be stored in the CRS. Some of the following data cannot be obtained for all chemical substances and chemical groupings. For all chemical substances where the data can be obtained, the following are stored:

1) EPA Chemical Registry Name Source (DE 20048:1) – the source of the EPA chemical registry name (e.g., EPA regulation name, EPA data system name, or other).

2) EPA Chemical Registry Name Context (DE 20052:1) – the name that identifies the circumstance in which that name has been used (e.g., iron and steel industry, pulp and paper industry, region of the country).

3) Molecular Formula (DE 5812:1) – the code that represents the number of atoms of each element in a molecule of a chemical substance. A complete code is required for all single, fully-defined chemical substances, and a partial code for mixtures where one or more components can be defined.

4) Chemical Substance Formula Weight (DE 5814:1) – the sum of the atomic weights of constituent atoms in a molecule of a chemical substance. The formula weight is required where a complete molecular formula code exists for a chemical substance.

5) Chemical Substance Type (DE 5818:1) – a descriptive name for types of chemical substances. Required for regulatory chemical classes.

6) Chemical Substance Definition (DE 20055:1) – the text that provides clarification to the identity of a chemical substance. Required when needed to completely, uniquely define a chemical substance.

7) Chemical Substance Linear Structure (DE 26069:1) – the code that represents the connectivity of atoms in a molecule of a chemical substance as a linear formula, such as Simplified Molecular Input Line Entry System (SMILES). A complete code is required for all single, fully-defined chemical substances, and a partial code where one or more components can be defined.

8) Chemical Structure Graphical Diagram (DE 26070:1) – a graphical representation of a molecule of a chemical substance as a two or three dimensional diagram. A complete representation is required for all single, fully-defined chemical substances, and a partial representation where one or more components can be defined.

9) Chemical Substance Comment (DE 20056:1) – the text that provides additional information about a chemical substance.

10) Chemical Substance Synonym (DE 5806:1) – the name that is used as an alternative for representing a chemical substance.

11) Chemical Synonym Source (DE 26073:1) – the name of the source of an alternate name for a chemical substance.

12) Chemical Synonym Context (DE 20042:1) – the name that identifies the circumstance in which that name has been used (e.g., iron and steel industry, pulp and paper industry, region of the country).

13) Chemical Synonym Status (DE 26071:1) – the name that documents the correctness of a synonym for a specific chemical.

14) Chemical Substance Classification (DE 26072:1) – the name of a schema that classifies chemical substances according to structural similarities.

e. The 13 REI-designated national program systems and all newly created or reengineered EPA data systems must have the capability to access and use the four standard data elements described in Section 4. c. above.

f. The EPA Chemical Internal Tracking Number will be used for database management and to ensure data integrity within EPA. The internal EPA number will not be displayed from the CRS, and it will not be made available for public use.

g. All chemical substances and chemical groupings referred to in official EPA publications and Web sites, including scientific, technical, regulatory and general information

documents must be identified by the CAS Registry Number, where it exists or can be assigned.

In addition, all chemical substances and chemical groupings must be identified by both the Chemical Substance Systematic Name, and the EPA Registry Name (common name) where it exists or can be assigned. These two names must be referenced at least once in each official EPA publication and Web site, after which either name may be used. A reference to a name may include: listing both at the beginning of a publication, providing document footnotes or by providing a crosswalk as an appendix/attachment.

h. When no CAS Registry Number and Chemical Substance Systematic Name exists, or can be assigned, the document or publication must use the EPA Chemical Registry Name. When no EPA Registry Name exists, the CAS Number and Chemical Substance Systematic Name must be used.

i. In the absence of a designated EPA Chemical Registry name for identification of chemical substances and chemical groupings, the Chemical Substance Systematic Name shall be used in place of an EPA Chemical Registry name.

j. The EPA Internal Tracking Number and the EPA Chemical Identifier shall each be created with a check digit (i.e., a digit generated by a mathematical algorithm and incorporated into the identification number to enable detection of transposition, transcription, and transmission errors, thus providing validity to the number).

k. The CAS Registry Number shall be validated according to the CAS check digit algorithm to ensure accurate representation in the CR.

5. PROCESSING

a. EPA will maintain a central CRS that will contain verified information identifying each chemical substance. The CRS will be available on the Internet.

b. All CAS Registry Numbers, Chemical Substance Systematic Names, EPA Chemical Registry Names, EPA Chemical Internal Tracking Numbers, and EPA Chemical Identifiers used in Agency products must be obtained from and validated with the EPA Chemical Registry System.

c. Procedure for initial population of the EPA Chemical Registry Name:

1) Identification of the major regulations and key databases to be used as name sources.

2) Determination of the candidate names by "Popularity Tally."

   (a) All names now in use for those chemical substances or chemical groupings in any of a specified set of major regulatory lists and key databases will be compiled. A tally of how many times a substance is called by the various names that have been used for it. The name with the highest tally will be the presumptive candidate name. Candidate names, together with all other names in the compiled list for each substance, are passed forward for technical review.

   (b) If the TSCA Inventory and related OPPT databases and regulations are the only sources where a particular chemical substance is mentioned, an EPA Chemical Registry Name will not be created. If another EPA office later determines that it needs to regulate, monitor, or otherwise track such a chemical, use the procedure in the following subsection (5.c.) for selecting an EPA Chemical Registry name.

   (c) Chemical groupings created through EPA regulations will be assigned EPA Chemical Registry Names by the Chemical Nomenclature Committee.

3) Review of the candidate names.

The EPA Chemical Nomenclature Committee (hereafter referred to as the Nomenclature Committee) will review the candidate name for each substance to determine if it is chemically accurate, unambiguous, and unique. This group may, for any reason it finds compelling, reject the initial candidate name and substitute any other name in the compilation as long as it is chemically accurate, unambiguous, and unique. Names approved or substituted through this process are designated "pending publication."

This group must reject any name that is not chemically accurate, unambiguous, or unique. It shall also be guided by the additional desirable characteristics for an EPA Chemical Registry Name, placing whatever relative weight it deems appropriate on each of those additional characteristics in its decision making on each specific chemical substance or chemical grouping it reviews.

   a) Additional “preferred” characteristics can include the following: simple, easy-to-understand, adherence to a particular naming convention (e.g., all halogenated methanes being referred to in the form of either halomethane or methyl halide), systematic use of valence symbols or terminology

(e.g., use Roman numerals only or Arabic numerals only), and other characteristics that the technical committee might identify during the process. The name should not be a trade name, a code name, or some sort of abbreviation.

b) The Nomenclature Committee will prepare comprehensive criteria for the selection of EPA Chemical Registry Names that will reflect the policies discussed herein.

4) Review for comment

Names designated "pending publication" will be available for comment for a limited time (30 days has been suggested), and comments will be reviewed by the Nomenclature Committee, who will make a final determination regarding the EPA Chemical Registry Name in each case. These final names will become the "Official" EPA Chemical Registry names once this final determination is made.

d. Procedure for assigning an EPA Chemical Registry Name to a chemical substance or chemical grouping already in the CRS but for which the EPA Chemical Registry Name does not exist.

1) An office determines that no EPA Chemical Registry Name has been assigned.

Note: The only situation in which this should occur once EPA Chemical Registry Names have been assigned is if the only databases and regulations that include references to this chemical are the TSCA Inventory and related databases.

2) The office requests that an EPA Chemical Registry Name be assigned.

Options:

a) The requesting office suggests one or more names to the technical committee. The Nomenclature Committee reviews the name(s) and determines whether it is acceptable (i.e. chemically correct, unique, and unambiguous). If naming conventions are in place, those can be applied at this time.

b) The Nomenclature Committee selects a name.

Note: Option a) is probably preferable since the requesting office will already have worked on the chemical's identity and should have identified a name in common usage in the scientific or regulatory literature.

3) Review for comment. Since it is presumed that the requesting office wants to add the chemical to a database or regulation as soon as possible, this review period should be kept as short as possible (two weeks?). The reason a review period is needed is, if another office has been considering adding this chemical to a database, regulation, or list, it might have identified an alternate name for the chemical.

e. Any EPA office, data administrator, state, tribe, other stakeholder, or other Federal agency may request that a new chemical substance be added to the CR, using the following procedure:

1) Perform a search of the CRS based on name, name fragment, molecular formula, or other EPA data identifier (e.g., hazardous waste code, parameter code, pollutant code).

Whenever there are ambiguous search results and the unique identification of a substance is in question, contact the Data Steward for the CR for assistance.

2) Where a chemical substance has not been registered previously in the CR, submit a request to the Data Steward for the CR, with as much of the following data as is available: the CAS Registry Number, Chemical Substance Systematic Name, and other unique, unambiguous names used to identify the substance. The requestor shall also provide the rationale for registering the chemical substance or grouping.

3) After verifying that the submitted new chemical substance does not exist in the registry, the substance will be added to the Chemical Registry according to the procedure developed by the Data Steward for the Chemical Registry.

a) If the submission in 5.e.2 contains complete information equivalent to that needed by the Data Steward for the CR to assign the appropriate EPA chemical identifiers and add the substance or group, the substance shall be added as soon as possible, with a goal of completing the registration within 3 working days.

b) If the submittal is incomplete or needs a CAS Registry Number assigned, the appropriate EPA chemical identifiers will be assigned within a reasonable time, that time being dependent on the complexity of the review required by the Data Steward for the CR.

4) If the chemical substance is unique, the requesting office should submit recommendations for the EPA Chemical Registry Name (whether or not it has

been assigned a CAS Registry Number and CAS name). These should be accompanied by documentation showing that the name is in common usage. Names not selected will be designated as synonyms (as long as they have been used to describe the chemical in the literature and are accurate). The Nomenclature Committee will not attempt to identify additional synonyms for chemicals.

If a naming convention has been established for this particular type of chemical, that could also be used to establish the EPA Chemical Registry Name.

5) Review for comment. A review period, length to be designated, shall be provided for other offices to suggest alternate EPA Chemical Registry Names.

f. Procedure for changing the EPA Chemical Registry Name for a chemical substance contained in CRS that already has an EPA Chemical Registry Name.

1) This procedure should be implemented rarely. It is not clear what occasions might arise that would create a need for it. The Nomenclature Committee would be the preferred source for developing a procedure for changing the name. The following scenarios might create a need for that change:

a) An overlap of EPA Chemical Registry Names. In this instance, the name would no longer be unique.

b) Identification of an EPA Chemical Registry Name being used for another chemical, whether by another agency or in the scientific literature.

c) Identification of an EPA Chemical Registry Name no longer being chemically correct (e.g., if the chemical in question actually has a different structure or composition than that being described).

2) Procedure

a) The Nomenclature Committee may identify the problem on its own, or the problem may be identified for the technical committee by a program office, individual staff, or even the public.

b) The Nomenclature Committee shall review the problem and determine what steps need to be taken.

The proposed solution shall be reviewed by the program offices before being published for public comment by the technical committee.

3) Once a procedure for a particular type of problem has been established and approved, subsequent cases can be published by the Nomenclature Committee for comment by EPA offices or their partners.

6. ROLES AND RESPONSIBILITIES

a. The Chief Information Officer (CIO) will:

1) Ensure adherence to these business rules and will be responsible for resolving conflicts and issues relating to these business rules, including applicability.

2) Provide guidance and technical assistance to program offices and the regulated community in meeting the requirements of this standard.

3) Establish a permanent Chemical Identification Standard Standing Committee, consisting of participants from all EPA Offices, under the leadership of one designated lead office.

4) Appoint a Data Steward for the Chemical Registry, who will be responsible for the accuracy, reliability, and currency of the data content, and a Data Standard Steward for Chemical Substance Identification, who will be responsible for the accuracy, reliability, and currency of the standard.

5) Lead the effort to build a CRS that will include EPA's chemical identification data and ensure public access to it via the Internet.

6) Maintain a repository of metadata in the EDR of standard data elements and business rules for the Chemical Identification Data Standard.

7) The CIO will establish an agency process and operating procedures to review agency documents and publications for compliance with this standard.

8) Be responsible for issuing waivers from compliance with this standard.

b. Senior Information Resources Management Officers (SIRMO) and regional Information Resource Management (IRM) Branch Chiefs will:

1) Promote compliance with this standard.

2) Approve application for waiver from this standard and submit it to the CIO, who is responsible for final disposition of the application.

3) Appoint one or more representatives to the Chemical Identification Standard Standing Committee, one of which shall be designated a voting member who is authorized to speak for the Office.

4) Ensure that new regulations, documents, and automated systems identify chemicals according to this standard.

c. National System Program Managers responsible for the Reinventing Environmental Information (REI)-designated national program systems will:

1) Ensure that this standard is implemented as applicable in their systems.

2) Work collaboratively with the CIO on continuing standards development and implementation.

3) Identify and bring forward potential conflicts between these business rules, the underlying standards, and program systems needs.

d. The Chemical Identification Standard Standing Committee will be responsible for providing management oversight for Chemical Identification.

e. The EPA Chemical Nomenclature Committee, referred to as the Nomenclature Committee, shall be responsible for assigning EPA Chemical Registry Names, as appropriate, to chemical substances and groups of concern to EPA.

e. The Data Steward for the CR, or an authorized delegate, will be responsible for the following:

1) All communications with CAS, including searches of the CAS online database in connection with obtaining CAS Registry Numbers and Chemical Substance Systematic Names for chemical substances registered in CR.

2) Work in collaboration with data stewards from states and national systems to develop and maintain the master chemical vocabulary (i.e., the unique identification data) in the Chemical Registry.

3) Ensure the timely addition of new chemicals to the Chemical Registry.

4) Appoint and chair the EPA Chemical Nomenclature Committee.

f. The Data Standard Steward for Chemical Substance Identification will be responsible for the following:

1) Evaluating the data standard, including standard data elements and business rules, at no more than 5-year intervals, and making recommendations for changes as needed.

2) Ensuring that the current standard and all previous versions of this standard will be available in the EDR.

7. IMPLEMENTATION

a. The Agency will provide support functions for the development of a CRS where identification information about chemical substances will be stored and an EPA Internal Tracking Number will be created. There will be continual maintenance of the support functions.

b. Guidance, accessible data values, and appropriate tools to facilitate access to the CRS database will be provided by the second quarter of Fiscal Year 2001 (FY01).

c. Initial population of EPA Chemical Registry Names in the Chemical Registry will be completed by the end of the second quarter of Fiscal Year 2002 (FY02).

d. The EPA Chemical Registry Name shall be implemented over a transition period of one year or more. Program offices shall include the name currently being used, in parallel with the designated EPA Chemical Registry Name in regulations and published documents.

e. EPA's national systems, as identified by the REI program, will implement the Chemical Substance Identification Data Standard, accepting new data as specified by the standard no later than the second quarter FY03.

8. PROVISION FOR WAIVER

a. The Agency's CIO may grant waivers for sufficient reasons.

b. The EPA policy shall be followed with respect to the application for and granting of waivers.

9. MAINTENANCE

a. The standard and these business rules shall be reviewed by the Data Standard Steward for Chemical Substance Identification in one year after issuance of this standard and at annual intervals until the year 2004. Subsequently the standard will be reviewed at no more than 5-year intervals, and updates recommended as necessary.

b. The Data Standard Steward for Chemical Substance Identification shall ensure that the current standard and all previous versions of the standard be available in the EDR.

c. The Data Steward for the CR, in cooperation with the Chemical Identification Standard Standing Committee, shall be responsible for the ongoing maintenance and update of the central registry of chemical identification data.

10. REFERENCES

a. *Interim Data Standard for Chemical Identification*, Tom Maloney, May 6, 1999.

b. *Reinventing Environmental Information (REI) Interim Chemical Information Data Standard*, Chemical Data Standard Work Group Report, April 22, 1999.

c. *Summary Report of Standard Data Elements for Chemical Substances*, SDC-0055-057-TC-7015, December 2, 1997.

d. *Rules for Representation of Chemical Data in Envirofacts Pilot Master Chemical Integrator*, SDC-0055-057-LF-3019A, June 10, 1994.

e. *Chemical Abstract Service Registry Number Data Standard*, IRM Policy, Standards and Guidance/IRM Strategic Planning Documents, EPA Directive No. 2180.1, June 26, 1987.

f. *EPA Chemical Registry Name Selection Procedures*, Drafted by the Chemical Name Selection Subgroup of the Chemical Identification Standard Business Rules Workgroup, Draft 1.1, February 29, 2000 (rev. April 9, 2000).